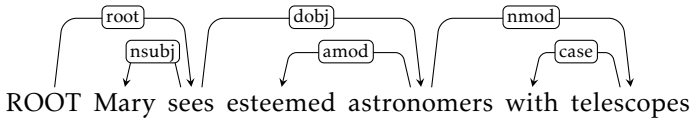
Dependency Parsing exercises: Transition-based arc-standard parsing

Deadline: 17.05.2021 for Ex. 1, 20.05.2021 for Ex. 2

Please send completed solutions to waszczuk@hhu.de and evang@hhu.de with subject "dependency homework" and attachment "ex5_lastname(s).pdf".

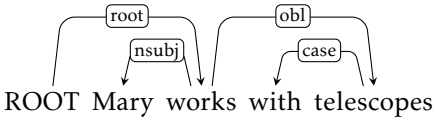
1. Enumerate the configurations and transitions an arc-standard transition-based parser goes through when parsing the sentence:[1]

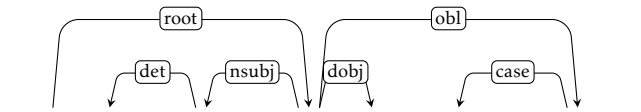


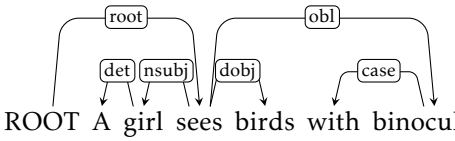
   A transition is a left-arc, right-arc, or shift operation (LA, RA, or SH). At each step, indicate the operation, the contents of the stack, the input buffer, and which dependency is added, if any:

TRANSITION	STACK	BUFFER	ARCS
	[ROOT]	[Mary sees esteemed astronomers with telescopes]	$\emptyset$
SH	[ROOT Mary]	[sees esteemed astronomers with telescopes]	
LA_{SBJ}	[ROOT]	[sees esteemed astronomers with telescopes]	$+(\text{Mary} \overset{\text{SBJ}}{\leftarrow} \text{sees})$
...	...	...	...

2. It turns out that we have a small corpus with some more information on the kinds of attachments "sees" and "telescopes" tend to have:







   In this corpus, all ~~prepositions~~ prepositional phrases are attached to verbs! Assume that an arc-standard model is trained with the following feature templates:

[1] That is, the configuration/transition sequence which yields the corresponding dependency tree. If there are several such sequences, it is enough to describe one of them.

- the word below the top of the stack
- the word on top of the stack
- the first word in the input buffer
- the second word in the input buffer

An example of a configuration and its features would look as follows:

- Configuration: ([ROOT], [Mary works with telescopes], $\emptyset$)
- Features: (ε, ROOT, Mary, works)

Here's another example (the features do not rely on the set of arcs, hence the . . .):

- Configuration: ([ROOT sees esteemed], [astronomers with telescopes], . . .)
- Features: (sees, esteemed, astronomers, with)

(a) We expect to arrive at a different analysis of the sentence in Ex. 1 if we train on this corpus. Assume we train a simple nearest neighbor classifier, i.e.:

- For a given parsing configuration, the parser chooses the transition of the single most similar configuration in the training dataset.[2] If there are several candidates, the majority class should be chosen.

Which configuration, transition, and features in the training dataset will be responsible for changing the attachment of the ~~preposition~~ prepositional phrase as compared to the analysis in the previous exercise?

(b) Suppose we have a larger corpus about telescopes, and it turns out there is a lot of ambiguity with respect to telescopes and their attachments. What kind of feature would work best to deal with ambiguity: part of speech tags, lemmas, or word forms? What are the trade offs of different feature kinds?

[2] To measure the similarity between two configurations, count the number of feature values they share. For instance, $x = (\varepsilon, \text{ROOT}, \text{Mary}, \text{works})$ is more similar to $y = (\varepsilon, \text{ROOT}, \text{John}, \text{works})$ than to $z = (\text{ROOT}, \text{Mary}, \text{works}, \text{with})$ because x and y share the two top elements of the stack (ε and ROOT) and the second word in the buffer (works), while x and z share no common features. Note that while both x and z have the feature value ROOT, in x it corresponds to the top of the stack and in z it corresponds to the word below the top of the stack.